Confidence Regions for Evolutionary Trajectories

Charles E. McCulloch, Michael D. Boudreau[1]

Biometrics Unit and Statistics Center, Cornell University, Ithaca, NY, 14853, U.S.A.

and

Sara Via

Department of Entomology and Section of Ecology and Systematics
Cornell University, Ithaca, NY, U.S.A. 14853

BU-1178-MC February 1995

[1] 33 Knowles St., Pennington, NJ 08534

Summary

We derive confidence regions for the evolutionary trajectories derived from a model of Via and Lande (1985, *Evolution* **39**: 505-522). We utilize a nested, parametric bootstrap to calculate the confidence regions and show that a likelihood-based approach provides an unsatisfactory solution. We calculate the predicted evolutionary trajectories and confidence regions using data from an experiment on pupal weight of *Tribolium* beetles in two different flour types. With realistic sample sizes for experiments that estimate genetic parameters in multiple environments, the confidence regions are likely to be large, potentially limiting the ability to test hypotheses about evolutionary trajectories. The problem may be less acute for trajectories of characters expressed in only a single environment, because sample sizes can generally be larger than in multi-environmental studies.

Key Words: bootstrap, likelihood ratio, variance components, mathematical models.

1. Introduction

Via and Lande (1985) developed a mathematical model describing the dynamics of evolutionary change in quantitative characters (i.e., variables) in a population of organisms inhabiting different environments. The model gives the change from one generation to the next in the mean phenotypic value of the characters in each environment as a function of certain population parameters, including the phenotypic and genotypic variance-covariance matrices, **P** and **G**. If the model is applied repeatedly over time, it traces a curve followed by the joint mean value of the characters in two or more environments, which is called an evolutionary trajectory (see Figure 1 for an example). The Via-Lande model assumes that **P** and **G** are known, whereas in practice they must be estimated and are subject to sampling variability. In this paper we show how to construct confidence regions for evolutionary trajectories via a nested parametric bootstrap and show that a likelihood-based approach gives an unsatisfactory solution. These confidence regions are necessary for summarizing the descriptive power of predictions made from estimated models and for assessing evolutionary predictions made from the model when fit to actual data from evolutionary selection experiments. We illustrate the methodology by formulating a predicted set of trajectories for the evolution of pupal weight using data from an experiment on *Tribolium* beetles reared in an environment that is heterogeneous with respect to flour type.

2. The Via-Lande Model

Let $\underset{nm\times 1}{\mathbf{z}}(t) = \left(\underset{n\times 1}{\mathbf{z}_1'}(t), \underset{n\times 1}{\mathbf{z}_2'}(t), \cdots, \underset{n\times 1}{\mathbf{z}_m'}(t) \right)'$ denote the values at time t of n characters in m environments, where $\mathbf{z}(t) = \mathbf{x}(t) + \mathbf{e}$, with $\mathbf{x}(t)$ the additive genetic effects and $\mathbf{e}$ the environmental effects. The individual entries in $\mathbf{z}(t)$ are phenotypic values, with the values of a trait expressed in different environments termed character states. We assume that

$$\mathbf{x}(t) \sim \mathcal{N}(\boldsymbol{\mu}_{\mathbf{z}(t)}, \mathbf{G})$$

and (1)

$$\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \mathbf{E})$$

are independent of $\mathbf{x}(t)$. We further assume $\mathbf{E} = \text{diag}\{\mathbf{E}_1, \mathbf{E}_2, \cdots \mathbf{E}_m\}$.

The phenotypic covariance matrix is

$$\mathbf{P} = \mathbf{G} + \mathbf{E} = \mathrm{var}[\mathbf{z}(t)]$$

with the component matrix in the ith environment given by

$$\mathbf{P}_i = \mathrm{Var}[\mathbf{z}_i(t)] \ .$$

The Via-Lande model predicts the change in the mean phenotypic values, $\Delta \boldsymbol{\mu}_{\mathbf{z}_i}(t) = \boldsymbol{\mu}_{\mathbf{z}_i}(t+1) - \boldsymbol{\mu}_{\mathbf{z}_i}(t)$, as a function of the current mean, $\boldsymbol{\mu}_{\mathbf{z}_i}(t)$. Natural selection is assumed to be of a stabilizing Gaussian form with the optimum in the ith environment, given by $\boldsymbol{\delta}_i$, and with width $\boldsymbol{\Omega}_i$. The mean fitness in the ith environment is thus

$$\mu_{w_{i(t)}} = \left(\frac{|\mathbf{P}_i|}{|\mathbf{P}_i+\boldsymbol{\Omega}_i|}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\mu}_{\mathbf{z}_i}(t) - \boldsymbol{\delta}_i\right)'(\mathbf{P}_i+\boldsymbol{\Omega}_i)^{-1}\left(\boldsymbol{\mu}_{\mathbf{z}_i}(t) - \boldsymbol{\delta}_i\right)\right\}, \tag{2}$$

with a mean fitness over all environments of $\bar{\mu}_{\mathbf{w}(t)} = \mathbf{q}'\boldsymbol{\mu}_{\mathbf{w}(t)}$, where q_i is the proportion of the infinite population of genotypes in the ith environment. Using $c_i(t)$ as a weight corresponding to the notion of hard selection in which populations in a patchwork of different environments are regulated globally, we define

$$\mathbf{C}(t) = \mathrm{diag}\{c_i(t)\} \otimes \mathbf{I}_n \ ,$$

where

$$c_i(t) = q_i \frac{\mu_{\mathbf{w}(t)}}{\bar{\mu}_{\mathbf{w}(t)}} \ .$$

We are now in a position to state the Via-Lande model; see their paper (Via and Lande, 1985) for details:

$$\Delta \boldsymbol{\mu}_{\mathbf{z}(t)} = \mathbf{G}\mathbf{C}(t)\left\{-(\mathbf{P}_i + \boldsymbol{\Omega}_i)^{-1}\left(\boldsymbol{\mu}_{\mathbf{z}_i}(t) - \boldsymbol{\delta}_i\right)\right\}_c \ , \tag{3}$$

where $\{\mathbf{A}_i\}_c$ denotes a matrix with submatrices $\mathbf{A}_i$ arranged in a column.

So, given $\boldsymbol{\mu}_{\mathbf{z}(0)}$ and values of the population parameters **G**, **P**, **q**, $\boldsymbol{\Omega}$ and $\boldsymbol{\delta}$, (3) can be applied repeatedly to trace out an evolutionary trajectory as in Figure 1.

3. Confidence Regions

To incorporate the effects of parameter estimation in the mathematical model (3) it is necessary to consider the sources of variation in trajectories observed in the laboratory. Assuming (3) to be

correct, variation in observed trajectories can be attributed to two sources: variation arising from sampling individuals subjected to selection and measurement error. In this paper we consider only the variation attributable to $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}_i$, though our inferential framework can straightforwardly incorporate other sources (e.g., variation in estimating $\boldsymbol{\mu}_{\mathbf{z}_i(0)}$). For many experiments the variation in $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}_i$ will be the major source.

Our interest therefore centers on iterates of the equation

$$\Delta \hat{\boldsymbol{\mu}}_{\mathbf{z}(t)} = \hat{\mathbf{G}}\hat{\mathbf{C}}(t) \left\{ -(\hat{\mathbf{P}}_i + \boldsymbol{\Omega}_i)^{-1} \left(\hat{\boldsymbol{\mu}}_{\mathbf{z}_i(t)} - \boldsymbol{\delta}_i \right) \right\}_c , \tag{4}$$

where $\hat{\mathbf{C}}(t)$ is $\mathbf{C}(t)$ with $\mathbf{P}_i$ replaced by $\hat{\mathbf{P}}_i$. Since $\mathbf{P}_i$ enters $\mathbf{C}(t)$ in a nonlinear fashion, even $\Delta \hat{\boldsymbol{\mu}}_{\mathbf{z}_i(1)}$ is a nonlinear function of $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}_i$. As equation (4) is iterated to trace out an evolutionary trajectory, the mean value of the character becomes an even more complicated function of $\hat{\mathbf{G}}$ and the $\hat{\mathbf{P}}_i$. This, along with the fact that the distributions of $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}_i$ themselves will usually be unknown (except for the simplest of balanced designs and most amenable types of estimators, the distributions of variance-components estimators are unknown), means that it is not feasible to work out the exact, small sample distribution of $\Delta \hat{\boldsymbol{\mu}}_{\mathbf{z}_i(t)}$. The delta method is a possible approach to the large sample distribution problem but given the difficult distributions associated with $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}_i$ and the unknown rate of approach to normality other methods of attack have advantages. For computational ease we used ANOVA-type estimates for the illustration in Section 4. For unbalanced data, ML or REML estimation would be preferred (Searle, Casella and McCulloch, 1992).

For the above reasons we used a parametric bootstrap with the percentile method (Efron, 1982) to evaluate confidence regions for evolutionary trajectories. Before describing those methods in detail we first establish some notation. Define a curve, $\mathbf{c}(\boldsymbol{\theta})$, as the set of points on an evolutionary trajectory, $\mathbf{p}(\boldsymbol{\theta}, t)$, traced out through time until a fixed generation T dependent on a vector of parameters $\boldsymbol{\theta}$:

$$\mathbf{c}(\boldsymbol{\theta}) = \{\mathbf{p}(\boldsymbol{\theta}, t): t = 1, 2, \cdots, T\} .$$

A confidence region for a curve, R_C, is a set of regions for the points at times $t = 1, 2, \cdots, T$, which we denote by $R_C(t)$, and a curve is said to fall in the region if and only if its point at time t falls in

$R_{\mathbf{c}}(t)$ for all $t = 1, 2, \cdots, T$:

$$\mathbf{c}(\boldsymbol{\theta}) \in R_{\mathbf{c}} \Leftrightarrow \mathbf{p}(\boldsymbol{\theta}, t) \in R_{\mathbf{c}}(t) \text{ for } t = 1, 2, \cdots, T .$$

In our application $\boldsymbol{\theta}$ will consist of the distinct elements of G and $\mathbf{P}_i$.

To apply the parametric bootstrap we need to select an appropriate statistic. A logical choice would seem to be $2\,\ell\log[L(\hat{\boldsymbol{\theta}} ; \mathbf{y}) \Big/ L(\boldsymbol{\theta} ; \mathbf{y})]$, where L is the likelihood under model (1). The usual large-sample approximation to $2\,\ell\log[L(\hat{\boldsymbol{\theta}} ; \mathbf{y}) \Big/ L(\boldsymbol{\theta} ; \mathbf{y})]$ would be to select a confidence region for $\boldsymbol{\theta}$ as

$$R_{\boldsymbol{\theta}} = \left\{ \boldsymbol{\theta} : 2\,\ell\log \frac{L(\hat{\boldsymbol{\theta}} ; \mathbf{y})}{L(\boldsymbol{\theta} ; \mathbf{y})} \leq \chi^2_{k,1-\alpha} \right\} \tag{5}$$

for an appropriate choice of k. A confidence region $R_{\mathbf{c}}$ for the curves would then be generated by selecting all the curves associated with $\boldsymbol{\theta}$ falling in the confidence region for $\boldsymbol{\theta}$:

$$R_{\mathbf{c}} = \{\mathbf{c}(\boldsymbol{\theta}) : \boldsymbol{\theta} \in R_{\boldsymbol{\theta}}\} .$$

A parametric bootstrap approach would refine the choice of the chi-square critical value but retain the same form as (5). In the next section we show that methods like (5) based on the likelihood do not yield confidence regions which are of a useful form for evolutionary trajectories. Essentially, the likelihood is sensitive to changes in parameter values which the trajectories are not, leading to an interspersion of trajectories which are and are not in the confidence region.

We therefore turned to statistics based directly on distances defined in the trajectory space in order to form our bootstrap. For example, plots of bootstrap replications for the illustration in Section 4 (see Figure 2) suggested that a Mahalanobis distance in the log $\boldsymbol{\mu}_{\mathbf{z}_i(t)}$ might be appropriate. That leads us to the following methodology:

1. Form an estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}$. Pick the number of bootstrap replications, B and the number of generations, T. Set $j = 1$.
2. Generate a bootstrap replication, $\mathbf{z}^*_j$, according to the distribution given by (1) with $\boldsymbol{\theta}$ replaced by $\hat{\boldsymbol{\theta}}$.
3. Form an estimate of $\boldsymbol{\theta}$, $\hat{\boldsymbol{\theta}}^*_j$, based on $\mathbf{z}^*_j$.
4. Calculate $\mathbf{c}(\hat{\boldsymbol{\theta}}^*_j) = \{\mathbf{p}(\hat{\boldsymbol{\theta}}^*_j, t)$: $t=1,2, \cdots T\}$, where $\mathbf{p}(\boldsymbol{\theta}, t) = \ell\log \boldsymbol{\mu}_{\mathbf{z}(t)}$.

5. Calculate the Mahalanobis distances $D_j^*(t) = \left(p(\hat{\theta}_j^*, t) - p(\hat{\theta}, t)\right)' \hat{\Sigma}_p^{-1} \left(p(\hat{\theta}_j^*, t) - p(\hat{\theta}, t)\right)$, where $\hat{\Sigma}_p$ is the estimated variance-covariance matrix of $p(\hat{\theta}, t)$.

6. If j = B stop; otherwise increase j by 1 and return to step 2.

7. Select $D_\alpha(t)$ t = 1, 2, $\cdots$, T so that the number of curves such that

$$D_j^*(t) \leq D_\alpha(t) \qquad t = 1, 2, \cdots, T$$

is equal to αB.

The regions $R_c(t) = \left\{p(\theta, t): [p(\hat{\theta}, t) - p(\theta, t)]' \hat{\Sigma}_p^{-1} [p(\hat{\theta}, t) - p(\theta, t)] \leq D_\alpha(t)\right\}$ then generate a confidence region R_c with approximate confidence coefficient $1-\alpha$. This is the analog of (5) with distance in the trajectory space replacing the log likelihood ratio and $D_\alpha(t)$ replacing the chi-square cutoff. Notes on the bootstrap procedure:

1. Other sources of variation could easily be included in steps 1 and 2 by expanding the definition of θ.

2. Other distance measures could be used in step 5 and, in fact, for situations in which the bootstrap points are not roughly elliptical, Mahalanobis distances in step 5 would be inappropriate. Other possible methodologies are Hall (1987), in which nonparametric density estimators are used or Liu (1990), in which simplical depths are used.

3. In step 5 we used a nested bootstrap to calculate $\hat{\Sigma}_p$. If the data at this stage were exactly multivariate normal, this would be an example of prepivoting, which would be expected to reduce the error in the approximation to the confidence coefficient. See Beran (1987, 1988) for details and more generally applicable methods.

4. Illustration

We illustrate both the Mahalanobis distance method described in Section 3 and the unsatisfactory nature of likelihood-based confidence regions using data from an experiment in which pupal weights of flour beetles (*Tribolium castaneum*) were measured in two different flours (environments); for a full analysis of the data see Via and Conner (1995). A subset of the data was selected to form a balanced half-sib breeding design with ten randomly selected sires each mated to two randomly selected dams with two progeny of each dam placed in each of two environments. The balance was desired to make the estimation of G and P_i straightforward.

The mean square mean product matrices for the MANOVA tables were simulated directly as pseudo-random Wishart matrices to avoid simulating the original data and to more directly calculate $\hat{\theta}_j^*$. Estimates were forced to be nonnegative definite by setting negative eigenvalues equal to zero. The estimates of $\mathbf{G}$ and $\mathbf{P}_i$ for the selected data set were

$$\hat{\mathbf{G}} = \begin{bmatrix} 4.60 & 3.72 \\ 3.72 & 4.56 \end{bmatrix} \quad \text{and} \quad \hat{\mathbf{P}}_i = \begin{bmatrix} 6.84 & 1.14 \\ 1.14 & 4.02 \end{bmatrix}.$$

The experiment we conducted gave estimates of $\mathbf{G}$ and $\mathbf{P}_i$ but was terminated before evaluating selection. We therefore had no actual data on $\mu_{z(0)}$, Ω_i and δ_i but had to specify reasonable values in order to get predicted trajectories. $\mu_{z(0)}$ would ordinarily be taken as the initial mean values in environments 1 and 2. δ_i, the optimum, would have been chosen to be the mean pupal weight in the best environment (in this experiment – whole wheat with brewer's yeast). Ω_i was chosen to be diagonal since selection is assumed to occur independently in each environment. The actual values for the diagonal of Ω_i are difficult to specify; we chose approximately 40 $\mathbf{P}_i$.

Figure 3 shows the 95% confidence region for generation $t = 5$ (out of a total of $T = 10$ generations) and the values of 500 bootstrap samples using the Mahalanobis distance method described in Section 3. A nested bootstrap of 200 replications was used to determine $\hat{\Sigma}_p$. Figure 4 shows the confidence regions for a variety of generations.

The confidence regions in Figure 4 are disappointingly large in that predictions from the model are quite inaccurate. This compromises the utility of both model validation and prediction for models with $\mathbf{G}$ and $\mathbf{P}_i$ estimated at this level of accuracy. To investigate the effects of sample size, we increased the number of sires to 20, leaving all the other parameters the same. Figure 5 displays the results. The regions are somewhat smaller, but not appreciably so.

To use the likelihood approach with the half-sib design we must first make a minor modification to eliminate the nuisance parameters. Let Σ_S represent the variance-covariance matrix of the sire effects, Σ_D the variance-covariance matrix of the dam effects and Σ_E the error variance-covariance matrix. Then the trajectories only depend on $\mathbf{G} = 4\Sigma_S$ and $\mathbf{P}_i = \Sigma_{S,i} + \Sigma_{D,i} + \Sigma_{E,i}$, where the subscripts indicate those portions of Σ_S, Σ_D and Σ_E associated with the i^{th} environment. However,

the likelihood depends on Σ_S, Σ_D and Σ_E separately. Thus the appropriate modification of the likelihood ratio is to separate $\boldsymbol{\theta}' = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$ where $\boldsymbol{\theta}_1$ contains the distinct elements of Σ_S and $\Sigma_D + \Sigma_E$ and $\boldsymbol{\theta}_2$ those of Σ_D. We then form a confidence region using

$$R_{\boldsymbol{\theta}} = \left\{ \boldsymbol{\theta} : 2 \ell\text{og} \frac{L(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2; \mathbf{y})}{L(\boldsymbol{\theta}_1, \hat{\boldsymbol{\theta}}_2; \mathbf{y})} \leq \chi^2_{k,1-\alpha} \right\}. \tag{6}$$

For the *Tribolium* data we formed a region in this fashion. Figure 5 shows points (t=5, T=10) that are and are not in the region defined by (6), using $\alpha = .05$ and $\chi^2_{5,.95} = 11.1$. The points inside and outside the confidence region are completely interspersed, making the region calculated using the likelihood generally unsatisfactory for interpretations in the trajectory space.

5. Discussion

The rate and direction of phenotypic evolution is jointly determined by the phenotypic and genetic covariance structure of phenotypic traits and the intensity of natural selection (Via and Lande, 1985). Given a dynamic model of evolution that utilizes the covariances and selection intensities as parameters, evolutionary trajectories calculated for different combinations of parameter values can potentially be compared. This permits the evolutionary consequences of different parameter values to be evaluated. One of the most useful features of quantitative genetic models of evolution is that the parameters can be experimentally estimated on natural or laboratory populations using standard methods (e.g., Falconer 1991 for genetic parameters, and Lande and Arnold 1983 for a method to estimate the intensity of selection).

Using empirical estimates of the parameters of the models to formulate predicted evolutionary trajectories is of interest in an evolutionary context for several reasons. First, the descriptive power of evolutionary models like that of Via and Lande (1985) can be validated by establishing experimental populations in an appropriate selection regime and allowing them to experience selection for several generations. In such an experiment the evolutionary trajectory can be observed by measuring the average phenotype of the population each generation. This observed trajectory can then be compared to that which was predicted from a model with parameter estimates made from the experimental

populations before selection. This provides a method for validating the model. Second, the future evolutionary fate of particular contemporary populations may be of interest. For example, one could use an evolutionary model to predict the phenotypic evolution of captive populations of endangered species when subject to a particular selection regime, or to predict the rate of evolution of a species of pest insect that is exposed to a specified array of crop plants. Given validation of the descriptive power of an evolutionary model, the consequences of particular man-made changes on the future evolution of a given population can be evaluated through the formulation of predicted trajectories based on parameter estimates from the population in question.

However, in order to compare an observed evolutionary trajectory with one predicted from parameter estimates, or to evaluate the probability that a particular trajectory will occur, it is necessary to be able to estimate a confidence region around the predicted trajectory. Here we have compared methods for formulating such a confidence region.

We utilized a parametric bootstrap approach, and found that a likelihood statistic was not useful because there was no clear delineation of points inside the confidence interval from points outside the region. However, the use of the Mahalanobis distance, a statistic based in the trajectory space, appears to be more promising. Even so, our results reveal that the confidence regions for the trajectories are quite large in comparison to the length of the trajectory. This shows that the trajectories are quite sensitive to the values used for **P** and **G** and that very large sample sizes will be needed to derive accurate predictions from this or similar models. Validation of the models will likewise require large sample sizes. Our method could be used to derive some insight into how large a sample size is required to produce a predicted trajectory with a given level of accuracy by performing the calculations using trial values.

It is not unexpected to find that large sample sizes are required in order to estimate the parameters of an evolutionary trajectory with a precision that is adequate for the formulation of a useful prediction. There has recently been considerable discussion in the evolutionary literature about the imprecision of estimates of genetic variability (e.g., Shaw, 1991). It may be that the logistical difficulties of parameter estimation will limit the formulation of specific quantitative predictions of

the rate and direction of short-term evolution. Our results support this concern by revealing that predicted trajectories are very sensitive to the precision of the parameter estimates. Thus, a considerable effort will be required in order to accurately predict the course of evolution over even a few generations. This will be especially true for studies of evolution in heterogeneous environments, because the total sample size must be divided among environments, considerably reducing the precision of genetic estimates in each environment. The difficulty of obtaining a precise estimate of genetic parameters at even a single time compounds the additional problems with precision of predicted evolutionary trajectories that can be caused if the parameters change during the course of evolution (e.g., Turelli, 1988). However, the extent to which genetic covariances change through time is controversial (Kohn and Atchley, 1988; Wilkinson *et al.*, 1990), partially because of additional statistical difficulties in comparing covariance matrices (Cowley and Atchley, 1992; Shaw, 1992).

ACKNOWLEDGEMENTS

This research was partially supported by USDA Hatch 151-403 to Charles E. McCulloch and a Searle Scholar's award to Sara Via.

REFERENCES

Aptech Systems (1990). *Gauss 2.1 Users' Manuals.* Kent, Washington.

Beran, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74**: 457-468.

Beran, R. (1988). Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* **83**: 687-697.

Cowley, D.E., and W.R. Atchley. (1992). Comparison of quantitative genetic parameters. *Evolution* **46**: 1965-1967.

Efron, B. (1982). *The Jackknife, the Bootstrap and other Resampling Plans.* SIAM, Philadelphia.

Falconer, D.S. (1989). *Introduction to Quantitative Genetics.* Longman Press. London.

Hall, P. (1987). On the bootstrap and likelihood-based confidence regions. *Biometrika* **74**: 481-493.

Kohn, L.A.P., and W.R. Atchley. (1988). How similar are genetic correlation structures? Data from mice and rats. *Evolution* **42**: 467-481.

Lande, R. (1979). Quantitative genetic analysis of multivariate evolution, applied to brain:body size allometry. *Evolution* **33**: 402-16.

Lande, R., and S.J. Arnold. (1983). The measurement of selection on correlated characters. *Evolution* **37**: 1210-1226.

Liu, R. (1990). On a notion of data depth based on random simplices. *Annals of Statistics* **18**: 405-414.

Searle, S.R., G. Casella, and C.E. McCulloch. (1992). *Variance Components*, Wiley, New York.

Shaw, R.G. (1991). The comparison of quantitative genetic parameters between populations. *Evolution* **45**: 143-151.

Shaw, R.G. (1992). Comparison of quantitative genetic parameters: reply to Cowley and Atchley. *Evolution* **46**: 1967-1969.

Turelli, M. (1988). Phenotypic evolution, constant covariances, and the maintenance of additive variance. *Evolution* **42**: 1342-1347.

Via, S., and J. Conner. (1995). Evolution in heterogeneous environments: genetic variability within and across different grains in *Tribolium*. *Heredity* **74**: 80-91.

Via, S. and R. Lande (1985). Genotype-environment interaction and the evolution of phenotypic plasticity. *Evolution* **39**: 505-522.

Wilkinson, G.S., K. Fowler, and L. Partridge. (1990). Resistance of genetic correlation structure to directional selection in *Drosophila melanogaster*. *Evolution* **44**: 1990-2003.

APPENDIX

Computing was performed using the matrix language GAUSS (Aptech Systems, 1990) on an IBM-PC compatible. Normal random numbers were generated using the built-in function RNDN. Details are available from the authors. The methods described in this paper are computationally intensive but are easily feasible on a fast PC. A run with 1000 bootstrap replications each with 200 nested bootstrap replications through 10 generations required about six hours on a 33-Mhz PC.

Figure Captions

Figure 1: An evolutionary trajectory of the joint evolution of the mean phenotype over 500 generations traced out by model (3) starting from (.15, .27) and nearing the optimum of (.40, .50).

Figure 2: Distribution of 500 bootstrap samples at generation t = 5.

Figure 3: Distribution of 500 bootstrap samples of $\ell \log \mu$ (in log mg $\times$ 100) at generation t = 5 and a 95% confidence region based on Mahalanobis distance.

Figure 4: Simultaneous 95% confidence regions for each of ten generations plotted with the evolutionary trajectory (in log mg $\times$ 100). The mean phenotypes for each of the ten generations are denoted with triangles and the estimate of **G** is based on ten sires.

Figure 5: Simultaneous 95% confidence regions for each of ten generations plotted with the evolutionary trajectory (in log mg $\times$ 100). The mean phenotypes for each of the ten generations are denoted with triangles and the estimate of **G** is based on 20 sires.

Figure 6: Parameter values inside and outside the 95% confidence region for generation 5 using the likelihood ratio method. Points "inside" the confidence region are denoted by 'o'; those "outside" the confidence region by '+'.

Figure 1: An evolutionary trajectory of 500 generations traced out by model (3) starting from (.15, .27) and nearing the optimum of (.40, .50).

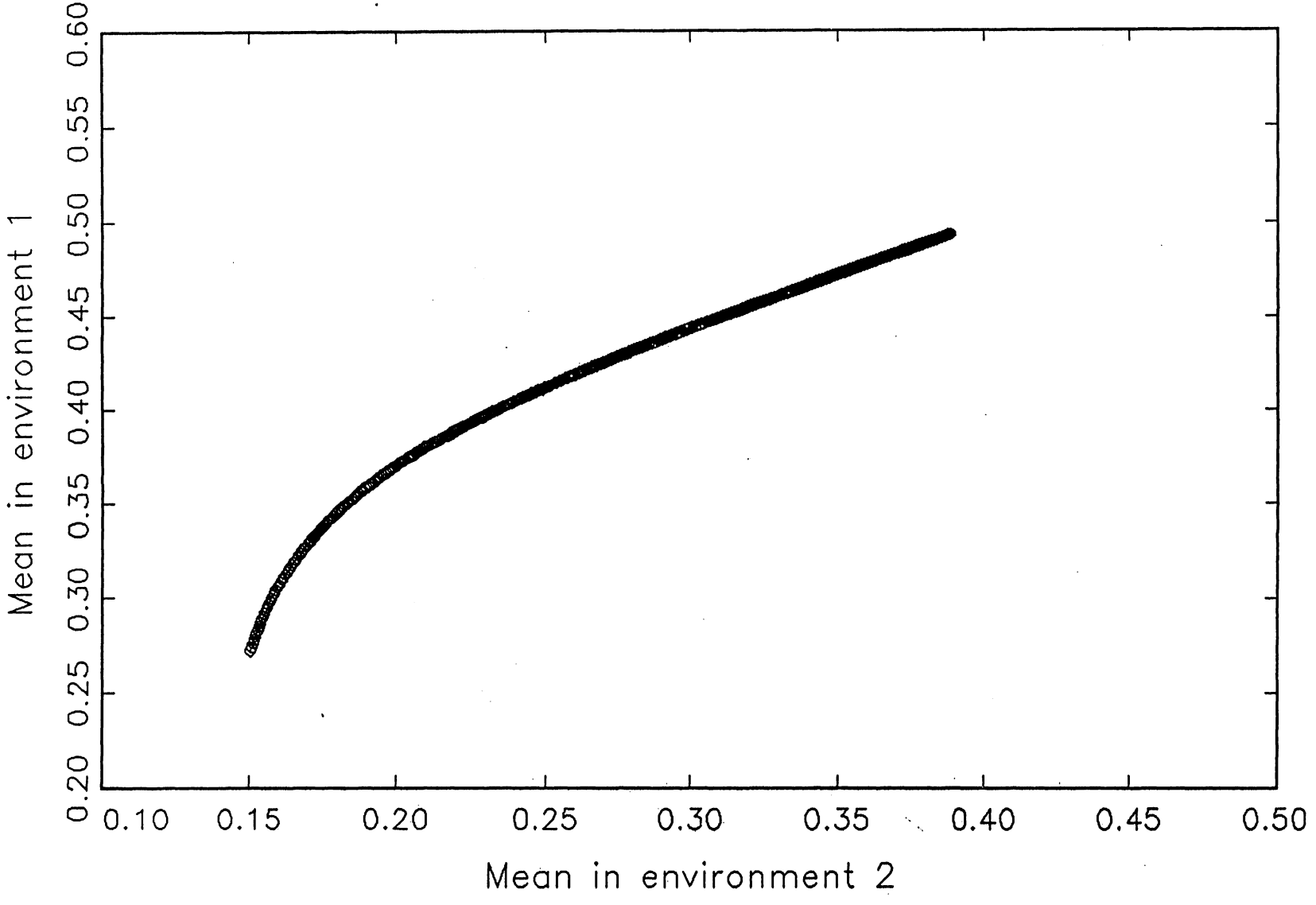

Figure 2: Distribution of 500 bootstrap samples at generation t = 5.

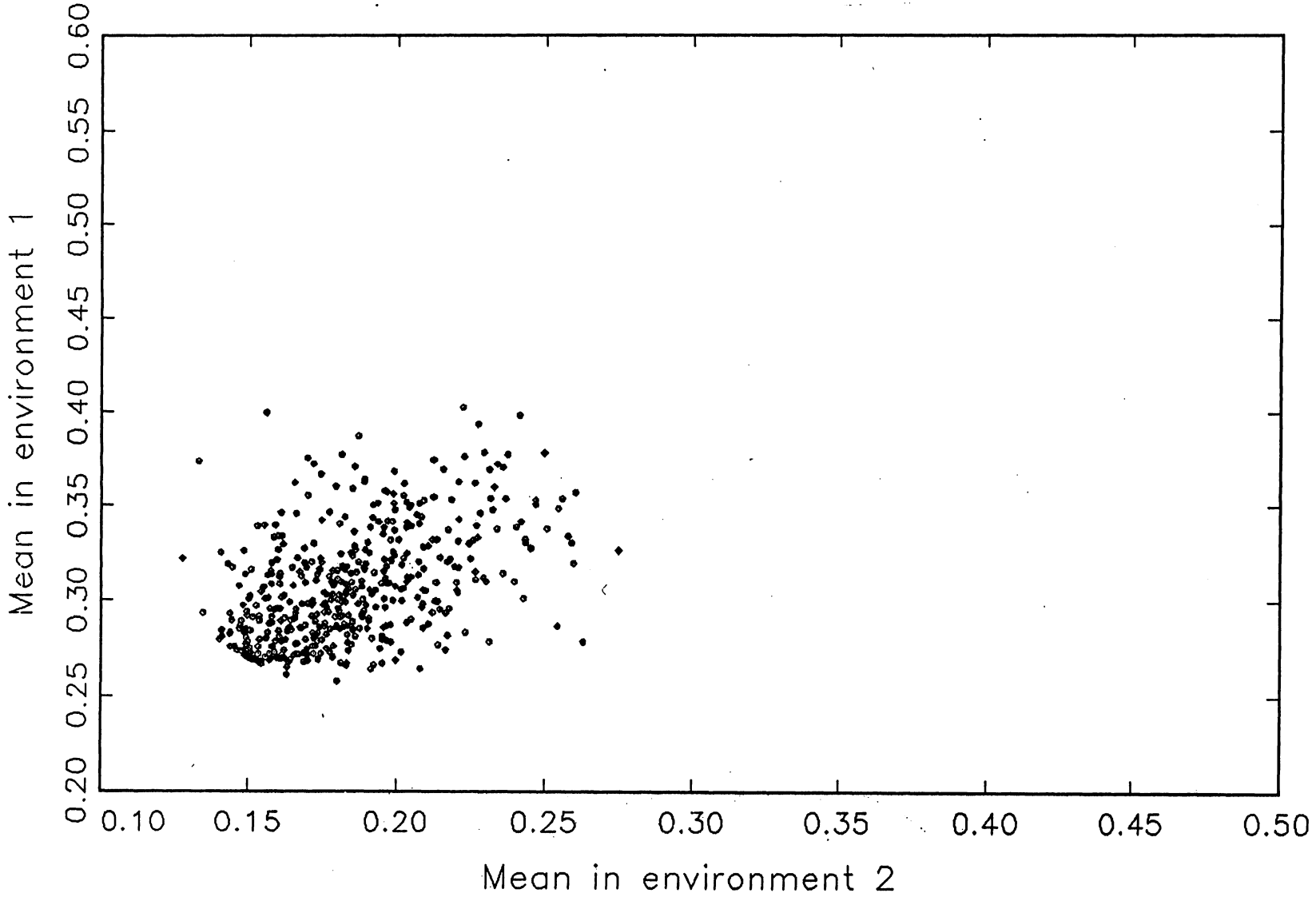

Figure 3: Distribution of 500 bootstrap samples of $\ell\log\mu$ (in log mg × 100) at generation t = 5 and a 95% confidence region based on Mahalanobis distance.

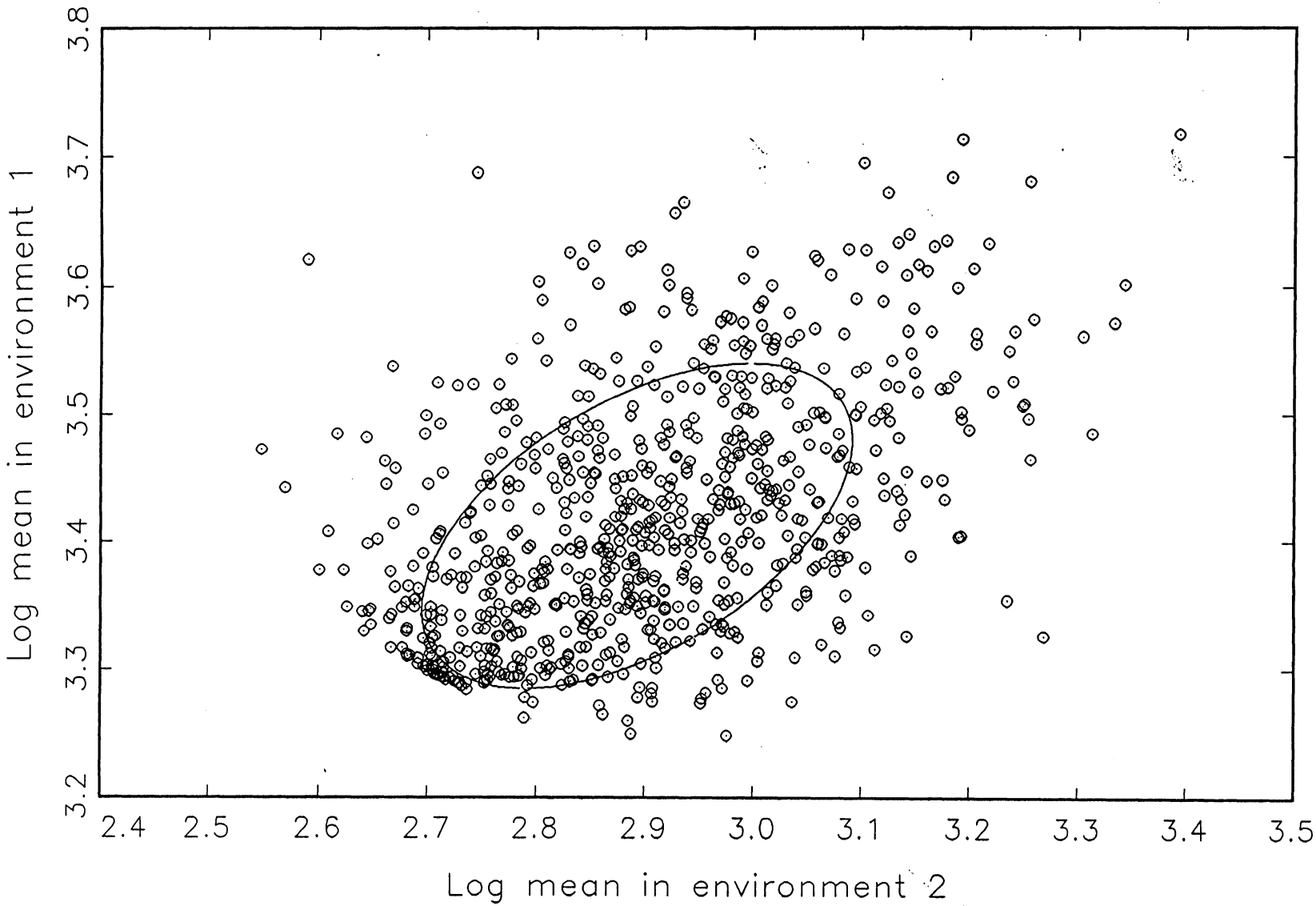

Figure 4: Simultaneous 95% confidence regions for each of ten generations plotted with the evolutionary trajectory (in log mg × 100). The mean phenotypes for each of the ten generations are denoted with triangles and the estimate of G is based on ten sires.

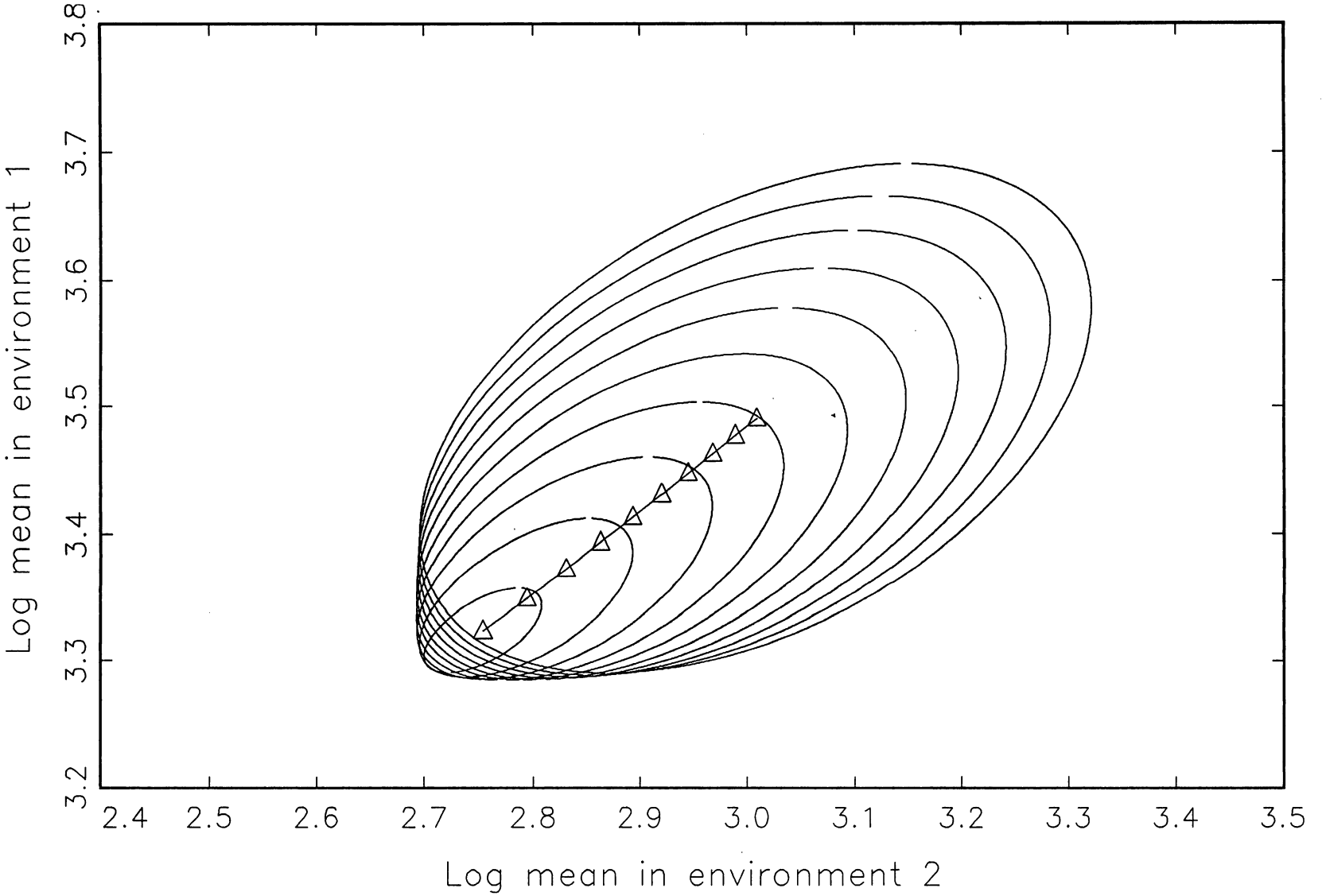

Figure 5: Simultaneous 95% confidence regions for each of ten generations plotted with the evolutionary trajectory (in log mg × 100). The mean phenotypes for each of the ten generations are denoted with triangles and the estimate of G is based on 20 sires.

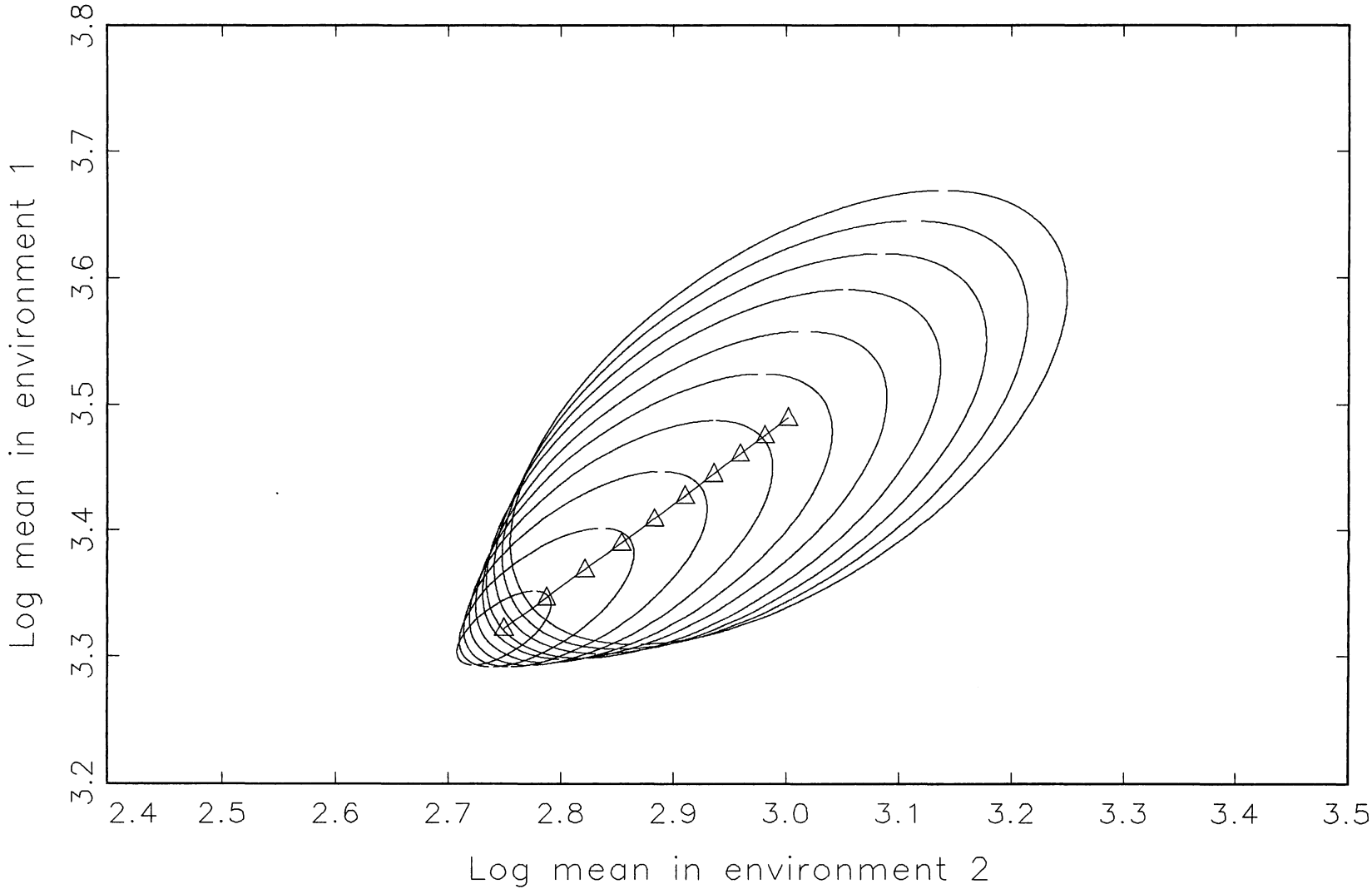

Figure 6: Parameter values inside and outside the 95% confidence region for generation 5 using the likelihood ratio method. Points "inside" the confidence region are denoted by 'o'; those "outside" the confidence region by '+'.